(* COUNTS *)

24

Expressing Counts

		Exh.	Page
Introduction	24-1	a	
- this chapter	24-2 to 24-3	b	
		1	24-8
24A. Relative importance of changes	24-3		
		c	
		2	24-11/12
24B. Expressions based on logarithms	24-4		
		d	
24C. Expressions based on square roots	24-7	e	
		3	24-15
		4	24-17/19
24D. Different grounds for choosing forms of expression	24-10	5	24-21/22
		6	24-23/24
24E. Examples involving minor grounds	24-14		
- symmetry	24-14 to 24-16	f	
- wide-stretched batches	24-16 to 24-20	7	24-26
- balance of spreads	24-20	8	24-27
		9	24-29
24F. Examples involving major grounds	24-20	10	24-30
- straight-line dependence	24-20 to 24-25	11	24-31
- additivity	24-25 to 24-32		
		g	
* * *		k	
		12	24-36/37
		13	24-38/42
24H. How far have we come?	24-33		
		l	
		14	24-47
* * *			
		p	
24K. How much does the choice of a free start matter? (background)	24-35	15	24-50
		16	24-51
		17	24-52
		18	24-53
24L. Roots and logs without tables (optional)	24-44	19	24-55
		20	24-56
* * *		21	24-57
24P. Problems for chapter 24	24-49		
* * *			
Table of started logarithms	24-8		
Table of double roots	24-11		

24

Expressing Counts

We are now ready to start on a new block of chapters -- a block distinguished by the nature of the responses rather than by the structures in which they appear. All the main types of structure we have so far dealt with will appear -- batches, y vs. x situations, two-way tables.

As a consequence we will meet again, perhaps in passing, all the main considerations that have already come up. Much of what we say will have somewhat of an air of repetition, but usually repetition with a slight difference or a new slant. Not only are we going to learn a little about quite new things, we are also going to get new breadth and depth on many that we have already touched. The refreshment and broadening of our insight and understanding we gain can -- and will -- be valuable -- both when our data is made up of counts and when it is not.

In particular, we will gain a clearer idea of how and why the analysis of data depends on the structure in which the given numbers are imbedded, a structure often visible in the data as it reaches us, but sometimes one we have to find and make both visible and useful. We can do this much better now that we have already met the three major kinds of structure and have a background against which we can compare and interrelate them.

* this chapter *

We often get our data by counting. Sometimes the total count is meaningful: How many planets revolve about the Sun? Or about other nearby stars? Here our attention centers on the count. In other situations what matters is a counted fraction: Out of a national sample of 1541 college students, 329 were found wearing beards. Here our attention centers on 329/1541 = .214 = 21.4%.

Counted rates -- as in the death rate is 12 per 1000 (per year), or he had an average of 0.13 mosquito bites per square inch of skin -- sometimes behave almost like counted fractions. More usually, however, counted rates behave rather like counts. After all, each count refers to something delimited, often in space or time -- to something with a "size". The cases where the "size" can be put in numbers and the count made into a counted rate -- by division by the corresponding size -- ought not be too different from those where this cannot be done.

What is most special about a count?

The roles of zero.

Zero is a bound beyond which we cannot go. In many sets of data zero is also a value that happens, perhaps often. When we are counting, we cannot have 1/2, 2 and 1/5, or 1/17. This matters most close to zero. Situations which might seem to deserve small fractions of different size cannot avoid all being recorded as "zero". The facts that there can be nothing between "one" and "two" and nothing between "two" and "three" have some

of the same effects, but these effects are usually small compared to the forcible compression at "zero".

24A. Relative importance of changes

When we have total counts, we usually want to choose a way of putting them into numbers that reflects what we really think about them. In this counts are like other responses.

What does "reflect what we really think about them" mean? It can mean many things, the simplest of which is merely that changes we regard as roughly equally "important" are expressed by roughly equal changes in the numbers that express our counts. Indeed, since our ability to compare the intuitive sizes of small changes is usually larger than our ability to make similar comparisons among large changes, we are likely to judge rough equality of change by using fairly small changes.

Consider two areas of land of roughly equal size -- we will call them counties for convenience. If one contains 10 wolves one year but 12 wolves the next, we would regard this change as a noticeable, but far from large, increase. If the other contains 1000 wolves one year but 1002 the next, we would regard this change as quite trivial. The fact that the difference in counts is exactly two wolves in both cases is true but quite misleading. (If you do not like to think about wolves, think about the same numbers for Boy Scout troops, Baltimore Orioles, airports, Presbyterian churches, parks, or cider mills.)

What if the second county has 1000 wolves one year, but 1200 the next? If these wolves come in large, irregular packs we may be tempted to feel that the change from 1000 to 1200 is no more important than the change from 10 to 12. If we feel this way, then we could usefully speak of a "20% change" in both cases.

We are, however, more likely to feel that a change from 1000 to 1200 is considerably "larger" than a change from 10 to 12. (Particularly if, at this season of the year, wolves run in pairs.) If this happens, whatever increase over 1000 that we judge to be about as "important" as a change from 10 to 12 will have to be a change to a count that is noticeably more than 1002 and noticeably less than 1200. In many cases, a change from 1000 to 1020, which falls in this range, is wisely judged to be about as "important" as a change from 10 to 12.

In the first case our intuition matches the use of "% change". In the second case it matches nothing so familiar. As we consider first a change from 10 to 12 and then a change from 1000 to 1020, the change in counts increases from 2 to 20 while the change in % decreases from 20% to 2%. In a sense, we want to go "halfway" between "number change" and "% change". As we saw long ago (chapter 3), this suggests using square roots.

24B. <u>Expressions based on logarithms</u>

Let us come to details, beginning with the situations where ratios tell the story.

To say that changes in % tell us what interests us most is to say that ratios are what matter. (Whether or not counts are in question, such a reaching for ratios seems to be, for example, the conventional and useful reaction to most sorts of economic data.) To say this -- and to have it roughly true -- is a long step forward, but it is no more than a single step: It lets us express changes, but our original task was to express the counts themselves -- and to do this in such a way that differences among our numbers will express, reasonably closely, the intuitive importance of changes in count.

Again we can turn to logs. Equal %s of change mean equal ratios -- before/after -- and hence equal logs of these ratios. Since the log of a ratio is the difference of the logs, equal logs of ratios means equal differences of logs.

Specifically, in our first wolf example:

$$\begin{aligned} \log 12 &= 1.07918 \\ \log 10 &= \underline{1.00000} \\ \text{difference} &= .07918 \end{aligned}$$

$$\begin{aligned} \log 1200 &= 3.07918 \\ \log 1000 &= \underline{3.00000} \\ \text{difference} &= .07918 \end{aligned}$$

so that, in these circumstances, expressing counts by their logs can give results corresponding to our intuition.

Only one difficulty needs a little attention. What if the count of wolves falls to 0? The log of zero is not a

useful number. We need to do something to avoid this kind of trouble. The easiest thing to do seems to be to express our count, not as a raw log, but rather as a

started log

-- by a number of the form

log (total count + free start)

where "free start" is a reasonably chosen small number, the same small number being used for all total counts that arise in a specific problem or a specific area.

Such choices as:

free start =

free start =

free start =

free start =

free start = 10

are each sometimes useful. In the absence of other guidance, using "free start = 0.25" when counts are "well behaved" but "free start = 1" when counts are "irregular" is likely to work fairly well.

Using such a free start distorts our differences a little, but usually not enough to make us uncomfortable. In the wolf example, for instance, the change from 10 to 12 now gives the same numerical difference as a change from 1000 to 1182. (I doubt if any of our intuitions is finely enough tuned to tell which is more closely matched to the change from 10 to 12, the change from 1000 to 1182 or the change from 1000 to 1200.)

Exhibit 1 gives -- in one way or another -- two-decimal values for any or all of these choices of "free start" and any count up to 99. Thus log (23 + 0.25) is given directly as 1.37, while log (51 + 1) is given as the sum of 1.71 (the value for 51 and "free start = 0.25") and 0.01 (for the *), namely 1.72.

For counts above 99, the break table of two-decimal logarithms (exhibit 3 of chapter 3, on page 3-10) should be used.

The expressions we have been looking at are based on logarithms, but they are not quite the same as unstarted logs. The difference is largest for small counts, becoming quite negligible for large ones.

While logs -- as we saw in chapter 3 -- play the role of a zeroth power, the expressions we have just met play the role of a power somewhat greater than zero -- greater by an amount that depends somewhat on the size of counts with which we are working. For most purposes we need no more detailed idea than this of where

log (free start + count)

stands in the scale of powers.

(It may sometimes be helpful to recognize that, for counts nearly equal to the free start, this form of expression behaves much like the square root of the count. More details will appear, for background, in section 24K.)

24C. Expressions based on square roots

So much, for the moment, for what we may reasonably do when equal ratios reflect equally important changes. What

exhibit 1 of chapter 24 -- tables

Started logs for counts

(For counts of 100 or more use exhibit 3 of chapter 3 on page 3-10)

A) TABLE, from 0 to 99, of log (count + o.25)

tens digit	units digit of count									
	0	1	2	3	4	5	6	7	8	9
0	-.60	0.10	0.35	0.51	0.63	0.72	0.80	0.86	0.92	0.97
1	1.01	1.05	1.09	1.12	1.15	1.18	1.21	1.24	1.26	1.28
2	1.31	1.33	1.35	1.37	1.38	1.40	1.42	1.44	1.45	1.47
3	1.48*	1.49*	1.51*	1.52*	1.53*	1.55*	1.56*	1.57*	1.58*	1.59*
4	1.60*	1.61*	1.62*	1.63*	1.64*	1.65*	1.66*	1.67*	1.68*	1.69*
5	1.70*	1.71*	1.72*	1.73	1.73*	1.74*	1.75*	1.76*	1.77	1.77*
6	1.78*	1.79	1.79*	1.80*	1.81	1.81*	1.82*	1.83	1.83*	1.84*
7	1.85	1.85*	1.86	1.86*	1.87	1.88*	1.88*	1.89	1.89*	1.90
8	1.90*	1.91	1.92	1.92	1.93	1.93	1.94	1.94	1.95	1.95
9	1.96	1.96	1.96*	1.97	1.97*	1.98	1.98*	1.99	1.99*	2.00

Note: The *'s are used in connection with B below

B) TABLE, from 0 to 29, of log (count + 1) For counts of 30 to 99 use panel (A), increasing *ed values by 0.01.)

	0	1	2	3	4	5	6	7	8	9
0	0.00	0.30	0.48	0.60	0.70	0.78	0.85	0.90	0.95	1.00
1	1.04	1.08	1.11	1.15	1.18	1.20	1.23	1.25	1.28	1.30
2	1.32	1.24	1.36	1.38	1.40	1.41	1.43	1.45	1.46	1.48

C) TABLE, from 0 to 29, of log (count + 0.1) (For counts of 30 to 99 use panel (A), without change.)

	0	1	2	3	4	5	6	7	8	9
0	-1.00	0.04	0.32	0.49	0.61	0.71	0.79	0.85	0.91	0.96
1	1.00	1.05	1.08	1.12	1.15	1.18	1.21	1.23	1.26	1.28
2	1.30	1.32	1.34	1.36	1.38	1.40	1.42	1.43	1.45	1.46

D) NOTES
(1) The accuracy of this table is usually unnecessarily great.
(2) For log(count + 3) enter panel B with "count + 2".
(3) For log(count + 10) enter panel B with "count + 9".

E) EXAMPLES of use of panels (A) TO (C)

count	log(count+0.25)	log(count+1)	log(count+0.1)	log(count+3)	log(count+10)
27	1.44	1.45	1.43	1.48	1.57
43	1.63	1.64	1.63	1.66	1.72
72	1.86	1.86	1.86	1.87	1.91

are we to do when, as is noticeably more common, we find a change from 1000 to 1020 a reasonable match for a change from 10 to 12?

We have already noted that what we want to do is to go "halfway" to the use of ratios, which means "halfway" to the use of logarithms. In dealing with the scale of powers, as in exhibit 19 of chapter 3, we saw that logs behave like zero powers, so that "halfway" from count to log count is likely to be the 0.5 power, corresponding to using $\sqrt{\text{count}}$. So let us try square roots.

For our wolf numbers we have:

$$\begin{aligned} \sqrt{12} &= 3.4641 \\ \sqrt{10} &= \underline{3.1623} \\ \text{difference} &= 0.3018 \end{aligned}$$

$$\begin{aligned} \sqrt{1020} &= 31.9374 \\ \sqrt{1000} &= \underline{31.6228} \\ \text{difference} &= 0.3146 \end{aligned}$$

The match is surely as close as our intuition can require.

We are thus free to use $\sqrt{\text{count}}$ -- or any fixed multiple or modification of it -- as a routine way of expressing counts. In practice we will rather often use just

$$\sqrt{\text{count}}$$

to be referred to as "root" or "simple root".

Sometimes we will want to use something a little different from twice this, namely

$\sqrt{2 + 4(\text{count})}$ for observed counts

$\sqrt{1 + 4(\text{count})}$ for fitted counts

to be referred to as

double root

or, where precision is essential, as started double root .

Exhibit 2 presents values of double roots -- to adequate working precision -- for counts of 0 to 199.

24D. Different grounds for choosing forms of expression

We have seen -- in past chapters -- at least four kinds of reasons for choosing one form of expression rather than another. Two of these were of lesser importance:

- near symmetry of pattern -- applies to individual batches.
- near equality of spread -- applies to two or more batches to be compared.

The other two were of major importance:

- simple apparent relationship of one variable to another -- in particular nearly straight-line relationship.
- near additivity in a two-way table -- as shown by a nearly horizontal diagnostic plot.

In earlier chapters we met these grounds for wise selection of a form of expression one after another, the order just used. We did notice that if a later ground applied, it should outweigh any or all earlier grounds.

We needed to say rather little about which ground dominated, however. So long as we are working with amounts,

exhibit 2 of chapter 24 -- table

Started double roots for counts

(observed or fitted)

A) DOUBLE ROOTS of OBSERVED COUNTS

tens digit	Units Digit of Count									
	0	1	2	3	4	5	6	7	8	9
0	1.4	2.4	3.2	3.7	4.2	4.7	5.1	5.5	5.8	6.2
1	6.5	6.8	7.1	7.3	7.6	7.9	8.1	8.4	8.6	8.8
2	9.1	9.3	9.5	9.7	9.9	10.1	10.3	10.5	10.7	10.9
3	11.0	11.2	11.4	11.6	11.7	11.9	12.1	12.2	12.4	12.6
4	12.7	12.9	13.0	13.2	13.3	13.5	13.6	13.8	13.9	14.1
5	14.2	14.4	14.5	14.6	14.8	14.9	15.0	15.2	15.3	15.4
6	15.6	15.7	15.8	15.9	16.1	16.2	16.3	16.4	16.6	16.7
7	16.8	16.9	17.0	17.1	17.3	17.4	17.4	17.6	17.7	17.8
8	17.9	18.1	18.2	18.3	18.4	18.5	18.6	18.7	18.8	18.9
9	19.0	19.1	19.2	19.3	19.4	19.5	19.6	19.7	19.8	19.9
10	20.0	20.1	20.2	20.3	20.4	20.5	20.6	20.7	20.8	20.9
11	21.0	21.1	21.2	21.3	21.4	21.5	21.6	21.7	21.8	21.9*
12	22.0*	22.0	22.1	22.2	22.3	22.4	22.5	22.6	22.7	22.8*
13	22.8	22.9	23.0	23.1	23.2	23.3	23.4*	23.5*	23.5	23.6
14	23.7	23.8	23.9	24.0*	24.0	24.1	24.2	24.3	24.4	24.4
15	24.5	24.6	24.7	24.8	24.9*	24.9	25.0	25.1	25.2	25.3*
16	25.3	25.4	25.4	25.6	25.7*	25.7	25.8	25.9	26.0*	26.0
17	26.1	26.2	26.3*	26.3	26.4	26.5	26.6	26.6	26.7	26.8
18	26.9	26.9	27.0	27.1	27.2*	27.2	27.3	27.4	27.5*	27.5
19	27.6	27.7	27.7	27.8	27.9	28.0	28.0*	28.1	28.2	28.2

Examples: Count = 27, expression = 12.2. Count = 165, expression = 25.7.

For *, see next panel.

(Cont'd)

exhibit 2 of chapter 24 (Cont'd)

B) DOUBLE ROOTS of FITTED COUNTS

	0	1	2	3	4	5	6	7	8	9
0	1.0	2.2	3.0	3.6	4.1	4.6	5.0	5.4	5.7	6.[illegible]
1	6.4	6.7	7.0	7.3	7.5	7.8	8.1	8.3	8.5	[illegible]
2	9.0	9.2	9.4	9.6	9.8	10.0	10.2	10.4	10.6	1[illegible]
3	11.0	11.2	11.4	11.5	11.7	11.9	12.0	12.2	12.4	12.[illegible]
4	12.7	12.8	13.0	13.2	13.3	13.5	13.6	13.7	13.9	1-.[illegible]
5	14.2	14.3	14.4	14.6	14.7	14.9	15.0	15.1	15.3	15.-
6	15.5	15.7	15.8	15.9	1 .0	16.2	16.3	16.4	16.5	16.[illegible]
7	16.8	16.9	17.0	17.1	17.2	17.3	17.4	17.6	17.7	1[illegible].[illegible]
8	17.9	18.0	18.1	18.2	18.4	18.6	18.6	18.7	18.8	18.[illegible]

For fitted counts $\geq$ 90, use panel A, subtracting 0.1 if there is a *.

we usually find relatively close agreement among these grounds as to what form of expression is favored.

The situation when we work with counts is a little different. The choice of expression is usually limited -- something based on a logarithm or something based on a square root. This is a narrow choice. Even grounds that we would ordinarily say agreed with one another quite closely can easily diverge to the extent that one favors a logarithm while the other favors a square root.

In the simplest and clearest situations, experience suggests that the minor grounds are likely to favor an expression rather like a square root. There are important caveats. If there are many zero counts, we can hardly hope for symmetry. If the fraction of zeroes varies sharply, we can hardly hope for equality of spread. If the counts we are treating as a batch vary because of a hidden influence -- perhaps swinging to and fro -- equality of spread is likely to come only when there is additivity in the imaginary two-way table where what we are comparing is one factor and the hidden influence is the other.

By contrast, the major grounds tend to favor a logarithm. The counted population of the U.S. before 1890 (see chapter 8) is an example for straight-line dependence; we shall see examples for additivity of effect shortly.

Roots and logs are not widely different, but the distinction seems to be consistent enough for us to want to bear in mind which is favored by which grounds for the choice of expression for counted responses.

Since the major reasons are the more important, it is natural for us to often try logs, raw or started, first. Even if additivity or straight-line dependence have no place in the data before us, more extensive or more structured data of the same kind may well offer them a place. Where this is so -- or even likely to be so -- it is probably good to start on some sort of a log scale.

There are situations, as we shall learn in a couple of chapters, where neither additivity or straight-line dependence have any apparent place. In these we will naturally turn to square-root expressions as our first choice.

24E. Examples involving minor grounds

* symmetry *

When there are no or very few zeroes, we ought to find three sorts of situation so far as symmetry goes, if our remarks above are correct:

- simple cases, where a square root works well
- transitional cases, which seem to fall between a square root and a log.
- extended cases, where a log seems most satisfactory.

When the ratio of largest count to smallest count is not very large, we may as well make a stem-and-leaf display and look at the five-point summaries. As part of a study of the distribution of plants of various species, Archibald gave the total counts of plants of all species in quadrats -- squares, that is -- of 20 square centimeters area in a Limonium salt marsh at Blakeney Point, Norfolk. Exhibit 3 sets out the stem-and-leaf and the five-point summaries.

exhibit 3 of chapter 24 -- Limonium marsh

Counts of total plants in 20 square centimeter quadrats, Limonium salt marsh, Blakeney Point, Norfolk

A) STEM-and-LEAF DISPLAY, SUMMARIZED

to	0	
4	0	5889
30	1	011111222233333333444444444
(26)	1	55555566667777788888899999
43	2	00000001111222223344444
20	2	5555566779
10	3	111234
4	3	559

20 + 26 + 43 = 99
hinges at 25

Hinges: 14 and 24

Median: 18

Extreme: 8 and 40

B) FIVE-POINT SUMMARIES for DIFFERENT EXPRESSIONS

	Raw	Root	Log
Extreme	8	2.8	.90
Range	14	3.7	1.15
Median	18	4.2	1.26
Range	24	4.9	1.38
Extreme	40	6.3	1.60
Median	18	4.2	1.26
Midrange	19	4.3	1.26
Midextreme	24	4.6	1.25

C) SOURCE

E.E.A. Archibald 1948. Plant populations I. A new application of Neyman's contagious distribution. Amals of Botany 12:221-235 Table -- at page 223.

> Expression in terms of logs seems to give the more symmetric result, but one should not give too much weight to how one batch seems -- or a few batches seem -- to behave.

* wide-stretched batches *

When the ratio of largest count to smallest count is large, and there are a reasonable number of counts, it is useful to sort out -- and count -- the counts in convenient groups according to equal steps in logs or roots or both.

Easy steps to use if the counts do not rise too high are shown in panel A of exhibit 4. Note that each end point appears in two intervals. The rule is to split the counts for each end point, throwing half into each interval. (Here, as elsewhere, the use of "h" for "and a half" makes appearances simpler and easier to grasp.) Panel B sets out Ashby's data on the number of plants of Salicornia Europea in 625-square-centimeter squares. Panel C illustrates the splitting rule in principle. Panel D shows how such calculations are most simply made.

Panel E shows the results for two sets of root intervals and one set of octaves. By using spacings for the two square-root summaries that bracket the spacing of the log summary we can avoid any serious distractions due to difference in spacing. This set of data seems much more symmetrical in roots than it does in logs. While this is still only one set of data, the ratio largest count/smallest count is big enough -- 66 -- as is the number of counts -- 100 -- to give us moderate confidence in this preference.

exhibit 4 of chapter 24 -- Salicornia quadrads

Tabulation By Roots and Octaves Illustrated on Numbers of Plants of Salicornia Europea

A) SOME HANDY STANDARD INTERVALS

Root	Log	Root
1-4	1-2	0-4
4-9	2-4	4-16
9-16	4-8	16-36
16-25	8-16	36-64
25-36	16-32	64-100
36-49	32-64	100-144
49-64	64-128	144-196
64-81	128-256	196-256
81-100	256-512	256-324
and so on	and so on	and so on

B) STEM-and-LEAF -- counts for 625-square-centimeter squares

to		
14	0	23334455556789
24	1	1122344778
36	2	002336888999
44	3	01146799
(8)	4	00235789
48	5	000011222345
36	6	011234578889
24	7	00334567889
13	8	228
10	9	116
7	10	04
5	11	07
3	12	7
2	13	23

(Cont'd)

exhibit 4 of chapter 24 (cont'd.)

C) DETAILS of TABULATION by OCTAVES -- "h" means "and a half"

sizes separated		counts	
2^k	other		
1		0 {0, 0}	h
2		1 {h, h}	
	3	3	4h
4		2 {1, 1}	
	5-7	6	7h
8		1 {h, h}	
	9-15	8	8h
16		0 {0, 0}	
	17-31	18	18
32		0 {0, 0}	

D) TABULATION of LOG INTERVALS

Summing down			Summing up		
To	Sum*	Diff	To	Sum*	Diff
256	0		1	0	
128	2	2	2	h	h
64	30h	28h	4	5	4h
32	61	30h	8	12h	7h
16	79	18	16	21	8h
8	87h	8h	32	39	18
4	95	7h	64	69h	30h
2	99h	4h	128	98	28h
1	100	h	256	100	2

*Half of values equal to each limit included.

Note: There are 2 up to 16, 79 = 100 - 21
There are 1 up to 8, 87h = 100 - 12h
There are 5 up to 4, 95 = 100 - 5
There are 0h up to 2, 99h = 100 - 0h

(cont'd.

exhibit 4 of chapter 24 (Cont'd)

E) ROOT-LOG-ROOT INTERVALS

1-4	5	1-2	0h		
4-9	8h	2-4	4h	16-36	18h
9-16	7h	4-8	7h	36-64	30
16-25	8	8-16	8h	64-100	24
25-36	11h	16-32	18	00-146	6h
36-49	11	32-64	30h		
41-64	18	64-128	28h		
64-81	17h	128-256	2		
81-100	6h				
100-121	3h				
121-144	3				
(root)		(log)		(root)	

F) SOURCE

E. Ashby 1935. The quantitative analysis of vegetation. Annals of Botany 49:779-802. Table III on page 790.

Exhibit 5 treats some counts of Singh and Das similarly. This time roots appear more unsymmetrical, while logs look relatively symmetrical.

* balance of spreads *

When we look at balance of spreads we tend to see the same three cases:

- where the square root serves
- transitional cases
- where the log is needed.

Exhibit 6 gives numbers of single, twin, and triplet births registered in Scotland for the years 1876 to 1895. Here roots give a somewhat better balance of spreads than do logs. If we had to compare only twins and triplet births, the preference for roots would be very clear. If, on the other hand, we had only to compare single and twin births the clear preference would be for logs.

This sort of transition from roots for smaller numbers to logs for larger ones is not uncommon. As panel E shows, a heavily started logarithm does quite well.

We will return to the examination of batches of counts which include zeroes -- counted or unknown -- in a later chapter.

24F. <u>Examples involving major grounds</u>

* straight-line dependence *

In chapter 8 we looked at a set of large counts -- the size of the population of the U.S. during the last century

exhibit 5 of chapter 24 -- Bornaya Infestation

Counts of plants of Bornaya brachiata for

quadrats of 2500 square centimeters area

A) STEM-and-LEAF

high	to		
232	17	0	44455555666778888
220	36	1	0122233446777899999
216	(15)	2	111233345567899
215	49	3	0224556789
195	3	4	1237
184	35	5	23478
176	30	6	11479
174	25	7	23446
172	20	8	0
161	19	9	1366
154	15	H	HHHHHHHHHHHHHHH
142	4		
123			
120			
105			

Down to	There are*	Down to	There are
225	1	256	0
196	4	128	12
169	9	64	27h
144	11	32	47
121	13	16	73h
100	15	8	85
81	19	4	98h
64	27h	2	100
49	35		
36	42h		
25	55		
16	73h		
9	83		
4	98h		
1	100		

*Including one-half the count of those equal to the limit.

(Cont'd)

exhibit 5 of chapter 24 (Cont'd)

B) ROOT-LOG-ROOT INTERVALS

1-4	1h	1-2	0	4-16	25
4-9	15h	2-4	1h	16-36	31
9-16	19h	4-8	13h	36-64	25
16-25	18h	8-16	11h	64-100	12h
25-36	12h	16-32	26h	100-144	4
36-49	7h	32-64	19h	144-196	7
49-64	7h	64-128	15h	196-256	4
81-100	4				
100-121	2				
121-144	2				
144-169	2				
129-196	5				
196-225	3				
225-256	1				
(root)		(log)		(root)	

C) SOURCE

B. N. Singh and K. Das 1938. Distribution of weed species on arable land. Journal of Ecology 26:455-466. Table II on page 461.

exhibit 6 of chapter 24 -- Scottish "twins"

Single, Twin and Triplet Births in Scotland,

1876-1895

A) THE DATA

	Single	Twin	Triplet
1876	126,534	1463	20
7	126,822	1429	12
8	126,773	1367	12
9	125,730	1452	20
0	124,570	1438	8
1881	126,171	1418	13
2	126,158	1415	7
3	124,458	1370	18
4	129,157	1457	19
5	126,100	1423	14
1886	127,890	1331	10
7	124,418	1335	18
8	123,275	1380	12
9	122,783	1423	10
0	121,526	1421	13
1891	125,986	1372	14
2	125,043	1468	17
33	127,110	1469	10
4	124,367	1450	13
5	126,486	1342	12

(Cont'd)

exhibit 6 of chapter 24 (Cont'd)

B) Cut STEM-and-LEAF DISPLAYS and FIVE-NUMBER SUMMARIES

single (unit 100)		twin (unit 10)		triplet (unit 1)	
121	5	13	0336	0	87
122	7	13	6787	1	223402036032
123	2	14	2311222	1	8987
124	5463	14	655665	2	20
125	790				
126	587116				
127	81				
128					
129	1				

Raw	Extreme	121,526	1331	7
	Hinge	124,438	1371	11
	Median	126,078	1422	13
	Hinge	126,654	1451	17.5
	Extreme	129,157	1469	20
Root	Extreme	348.6	36.6	2.6
	Hinge	352.8	27.0	3.3
	Median	355.1	37.7	3.6
	Hinge	355.9	38.1	4.2
	Extreme	359.4	38.3	4.5
Log	Extreme	5.085	3.124	.85
	Hinge	5.095	3.137	1.04
	Median	5.100	3.153	1.11
	Hinge	5.102	3.162	1.24
	Extreme	5.111	3.167	1.30

C) MEASURES OF SPREAD

		Raw	Root	Log
Diff of hinges	Single	2,216	3.1	.007
	Twin	80	1.1	.015
	Triplet	6.5	0.9	.20
Diff of extremes	Single	7,631	11.8	.026
	Twin	138	1.5	.043
	Triplet	13	1.9	.45

D) SOURCE

Annual Report of the Registrar General for Scotland for 1908. Table 7 at page ℓxvii.

E) RESULTS of using log (count + 200)

log (count) + (200)	Extreme	5.086	3.186	2.316
	Hinge	5.096	3.196	2.324
	Median	5.101	3.210	2.328
	Hinge	5.103	3.218	2.338
	Extreme	5.111	3.223	2.343
Diff of hinges		.007	.022	.014
Diff of extremes		.025	.037	.027

where log count gave a helpfully straight-line dependence. It is easy to find similar examples involving small counts.

* additivity *

Examples where additivity comes with logarithms are even easier to find.

In 1951, Neiswander reported counts on the white grub population in certain areas of turf. Eight years earlier the field in question had been divided into blocks, eight of which will be considered here. Each block was divided into plots, initially treated with different amounts of an insecticide -- lead arsenate. After the eight-year span, the amounts of insecticide remaining from each treatment, as shown by the number of grubs remaining, differed. Exhibit 7 gives the counts and their elementary analysis, which is rather obviously unsatisfactory.

The reader may verify that the corresponding diagnostic plot suggests trying logs. Since there are zeroes, the use of some sort of free start is almost essential. Since the data seem well-behaved, it is natural to begin by analyzing

log (0.25 + count)

as in exhibit 8. The residuals seem now to behave relatively well. The clarity has risen from 5.1 to 6.3. We have a better analysis.

In this example it took a rather sensitive look at the residuals to learn that a change of expression was needed. Equally, the gain from a change of expression, though worthwhile, was hardly striking. More striking examples are common.

exhibit 7 of chapter 24 -- residual arsenate

Neiswander's Counts of White Grubs in Turf Eight Years After Lead Arsenate
(Treatments: Pounds of lead arsenate per 1000 square feet)

Block	(Data in raw counts) Treatment 0	5	10	20	30	(Its elementary analysis) 0	5	10	20	30	eff	fit
1	16	5	1	0	1	1.3	-1.7	-0.8	-0.4	1.7	**-1.2**	7.0
2	10	3	1	1	0	-3.1	-2.1	0.8	2.2	2.3	**-2.8**	3.0
3	10	6	0	2	1	-3.9	0.1	-1.0	2.4	2.5	**-2.0**	3.8
4	7	6	0	0	0	-5.7	1.3	0.2	1.6	2.7	**-3.2**	2.6
5	14	5	6	4	1	-2.1	-3.1	2.8	2.2	0.3	**0.2**	6.0
6	26	10	6	2	0	7.1	-0.9	0.0	-2.6	-3.5	**3.0**	8.8
7	24	14	6	1	0	4.9	2.9	-0.2	-3.8	-3.7	**3.2**	9.0
8	20	14	4	3	1	1.5	3.5	-1.6	-1.2	-2.1	**2.6**	8.4
					eff	**10.1**	**2.1**	**-2.8**	**-4.2**	**-5.3**	**5.8**	0.0
					fit	15.9	7.9	3.0	1.6	0.5	0.0	-5.8

Clarity: 21.3/4.3 = 5.1

Source: Neiswander, C. R. 1951. Duration of the effectiveness of lead arsenate applied to turf for white grub control. J. Econ. Entomol. 44: 221-224

Problem: Calculate comparison values and make a diagnostic plot.

exhibit 8 of chapter 24 -- residual arsenate

Data of exhibit 7, expressed as log (0.25 + count)

Block	Data as 100 log (0.25 + count) 0	5	10	20	30	Elementary analysis--by means 0	5	10	20	30	eff	fit
1	121	72	10	-60	10	30	12	11	-50	- 9	-26	18
2	101	51	10	10	-60	- 9	-28	- 8	1	42	- 7	37
3	101	80	-60	35	10	19	- 3	-82	22	38	- 3	41
4	86	80	-60	-60	-60	21	46	-33	-24	15	-52	- 8
5	115	72	80	63	10	-27	-39	30	22	10	25	69
6	142	101	80	35	-60	9	- 1	39	3	-51	16	60
7	138	115	80	10	-60	7	15	41	-20	-49	14	58
8	131	115	64	51	10	-12	- 3	7	3	3	32	76
					eff	73	42	-19	-28	-69	44	00
					fit	117	86	25	16	-25		-44

low	high	to		
-82		3	L	LLL
-50		4	-4	9
-51		6	-3	93
		10	-2	7840
		11	-1	2
		17	-0	931389
		(7)	0	9771333
		16	1	925150
		10	2	122
		7	3	0098
		3	4	612
			11	

17 + 7 + 16 = 40, hinges at depth 10h

Hinges:	-16 and +20
Midspread:	36
Sides:	-52 and +56
Outside	1 and 0
(combined par)	(2)
Clarity:	$6.3 = \frac{142 + 84}{36}$

In the summer of 1950, C. B. Williams conducted trials of three light traps for catching insects. These trials were in a small piece of woodland (about 10 acres) on the famed Rothamsted Experimental Station (one of the first agricultural experiment stations in the world). Exhibit 9 sets out the broad outlines of the results.

(Since only one trap could be used at each location on a single night, it was necessary to have a detailed, balanced structure of traps and locations within three-day periods. The effects associated with these details were smaller than any of those in exhibit 9, and need not concern us for our present purpose.)

As the one cell of the elementary analysis not left to the reader shows, the two-way analysis of this data in raw counts is, to put it bluntly, pretty horrible.

Exhibit 10 shows the effect of using log (count + 1) which gives very much more satisfactory results. Indeed, if we rearrange the 15 row fits of exhibit 10 as a 5x3 table, we find the analysis shown in exhibit 11. Here, not only are the separate effects of Date-Group and Position moderately distinct, but we can see a clear indication of smaller residuals for position B.

exhibit 9 of ch. er 24 -- light traps

Williams's counts of Macrolepidoptera caught per night in each of three traps

Date-Group	Position	Counts			(Elementary Analysis*			
		Trap 1	Trap 2	Trap 3	1	2	3	Eff
16-18 Aug	A	17	100**	95				
"	B	40	28	95				
"	C	10	46	205				
19-21 Aug	A	24	195	137				
"	B	27	87	351				
"	C	20	247	1396	***			
22-24 Aug	A	42	356	482				
"	B	35	148	323				
"	C	18	216	421				
25-27 Aug	A	13	11	312				
"	B	24	115	156				
"	C	43	157	253				
28-30 Aug	A	10	22	109				
"	B	15	66	220				
"	C	31	188	671				

Notes:

* Elementary analysis is left to reader.

** Estimated.

*** The partial description based on row and column means gives a value of 551 + 31 - 165 = 417 here; thus the observed value of 20 has a residual of -397.

Source: Williams, C. B. 1951. Comparing the efficiency of insect traps. Bull. Entomolog. Res. 42: 513-517.

exhibit 10 of chapter 24 -- light traps

Williams's counts (from exhibit 9; expressed as log (count + 1) as per panel B of exhibit 1)

Date* group	Position	counts			Elementary Analysis--by means				
		Trap 1	Trap 2	Trap 3	1	2	3	eff	fit
16-18	A	1.25	2.00	1.98	+.03	+.18	-.22	-.17	1.75
"	B	1.61	1.46	1.98	+.47	-.30	-.16	-.23	1.69
"	C	1.04	1.67	2.31	-.11	-.08	.18	-.24	1.68
19-21	A	1.40	2.29	2.14	+.02	+.31	-.32	-.01	1.91
"	B	1.45	1.94	2.55	.00	-.11	+.12	.06	1.98
"	C	1.32	2.39	3.15	-.44	+.03	+.41	.37	2.29
22-24	A	1.63	2.55	2.68	-.13	-.19	.06	.37	2.39
"	B	1.56	2.17	2.51	.00	+.01	-.03	.17	2.09
"	C	1.28	2.34	2.62	-.34	+.18	+.02	.17	2.09
25-27	A	1.15	1.08	2.50	+.10	-.57	+.47	-.34	1.58
"	B	1.40	2.06	2.20	+.04	+.10	-.14	-.03	1.89
"	C	1.64	2.20	2.40	+.09	+.05	-.13	.16	2.08
28-30	A	1.04	1.36	2.04	+.09	-.19	+.11	-.44	1.48
"	B	1.20	1.83	2.34	+.01	-.08	+.07	-.10	1.82
"	C	1.50	2.27	2.83	-.18	-.01	+.18	.29	2.21
				eff	-.53	.07	.45	1.92	0.00
				fit	4.39	1.99	2.37	0.00	-1.92

* All dates are August 1950.

exhibit 11 of chapter 24 -- light traps

Two-way analysis of the 15 = 5×3 row effects from exhibit 8

A) ROW EFFECTS as a TWO-WAY TABLE

Date-group	Position		
	A	B	C
16-18	-.17	-.23	-.24
19-21	-.01	.06	.37
22-24	.37	.17	.17
25-27	-.34	-.03	.16
28-30	-.44	-.10	.29

B) ELEMENTARY ANALYSIS -- by means

				eff	fit
16-18	.16	.01	-.18	-.21	-.21
19-21	-.03	-.05	.08	.14	.14
22-24	.25	-.04	-.22	.24	.24
25-27	-.15	.07	.08	-.07	-.07
28-30	-.24	.01	.22	-.08	-.08
eff	-.12	-.03	.15	.00	.00
fit	-.12	-.03	.15	.00	.00

As Williams put it in the opening of his paper:

> "Our previous work on trapping insects by light, carried out for over eight years, had shown (Williams, 1937 and 1940) that the difference(s) between catches in different nights or in different traps are of a geometric nature, and not arithmetic or additative. Thus one trap catches on an average, say, twice as many insects per night as another trap--not 193 more insects per night.
>
> "Under these circumstances it is not correct to add catches obtained under different conditions and to use either the total or the arithmetic mean for the purpose of comparison. ...simplest... to convert the catch per night into a logarithm. ..."

When the structure of the data is such that we can compare the effectiveness of different forms of expression both clearly and sharply, the use of raw counts rarely proves better than the other choices we have illustrated.

Before closing this section, we ought to look back at section 24A and its emphasis on comparability of changes as a basis for choosing a form of expression. How is this to be interpreted in terms of major and minor grounds?

We can try to match comparability of changes to balance of spread. There will be some truth in such a match. We can do better, however, by matching it to additivity of effect. There will be even more truth in the latter match.

*****

24 . How far have we come?

This chapter has dealt with the effective expression of counts.

We have been led to using either square root or logarithmic expressions for counts, most usually of one of the started forms, such as the started double root,

$$\sqrt{2 + 4(\text{observed count})}$$
$$\sqrt{1 + 4(\text{fitted count})}$$

or the started logarithm

$$\log\ (\text{count} + \text{free start})$$

where "free start" is likely to be 0.25 or 1, but may be either larger or smaller.

We have begun to see the key role played by zero counts. Their pressure obviously enforces the use of a free start in logarithmic expressions and invisibly encourages the use of double roots. Their presence in appreciable number makes both symmetry and balance of spread unlikely.

When counts are small, irregularities tend to dominate differences in counts and square-roots are more likely to be useful. When counts are larger, effects--whether of recognized or unrecognized factors--play a larger role and logarithmic expressions are more likely to be useful.

Under the right circumstances, we can use any of the grounds for choosing expression that we have so far seen:

- symmetry of distribution.
- balance of spread.
- straight-line dependance.
- additivity of effect.

to decide how to express counts.

As always, each of the grounds is much more important than all the grounds that come before it in the list.

*****

24K. <u>How much does the choice of a free start matter?</u> (background)

We noted earlier that

$$\log(\text{free start} + \text{count})$$

behaves rather like

$$(\text{count})^{\text{power}}$$

with a power between zero and one that depends on the size of the count.

A useful rule of thumb says that this power is close to

$$\frac{\text{free start}}{\text{free start} + \text{count}}$$

Exhibit 12 shows how well this works for one simple case -- the batch consisting of three counts -- 0, 1, and 4. To make comparison easy, we have adjusted each form of expression -- multiplying by and adding constants -- to make their values 0 for 0 and 1 for 1. It is then easy to look at the value for 4 and see what power gives this value. In panel F the matching powers thus formed are compared with the mean value of the rule-of-thumb ratio -- the agreement is quite good.

If we have batches of more than three values, we cannot expect to match the same power to three of its lower values as to three of its higher values. The average value of the rule-of-thumb ratio can, however, be used to show us which power matches roughly and in some average sense.

Let us go on and look at the Neiswander data already analyzed in exhibits 7 and 8. Exhibit 13 compares various

exhibit 12 of chapter 24 -- casual example

Behavior of started logs of exhibit 1 for one

Particular Batch of Three Values

(0, 1, and 4)

A) VALUES of y TO VARIOUS POWERS

Power	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
y = 0	0	0	0	0	0	0	0	0	0	0
y = 1	1	1	1	1	1	1	1	1	1	1
= 4	1.15	1.32	1.52	1.74	2.00	2.30	2.64	3.03	3.64	4.00

B) BEHAVIOR of log (0.1+y)

	log (0.1+y)	1 + 0.96×(same)	
y = 0	-1.00	0	matching power 0.33
y = 1	0.04	1	
y = 4	0.61	1.59	

C) BEHAVIOR of log (0.25+y)

	log (0.25+y)	0.60 + 1.42×(same)	
y = 0	-0.60	0	matching power 0.46
y = 1	0.10	1.00	
y = 4	0.63	1.90	

D) BEHAVIOR of log (1+y)

	log (1+y)	3.33×(same)	
y = 0	0	0	matching power 0.61
y = 1	0.30	1.00	
y = 4	0.70	2.33	

E) BEHAVIOR of log (3+y)

	log (3+y)	-3.82 + 8×(same)	
y = 0	0.477	0	matching power 0.78
y = 1	0.602	1.00	
y = 4	0.845	2.94	

(Cont'd)

exhibit **12** of chapter 24 (Cont'd)

F) BEHAVIOR of (free start)/(value + free start)

free start	0.1	0.25	1	3
value = 0	1.00	1.00	1.00	1.00
value = 1	0.09	0.20	0.50	0.75
value = 4	0.02	0.06	0.20	0.43
mean	0.34	0.42	0.57	0.73
(matching power)	(0.33)	(0.46)	(0.61)	(0.78)

exhibit 13 of chapter 24 -- residual arsenate

Counts of white grubs analyzed with
log (free start + count) for various values of "free start"
(see exhibit 7)

A) FREE START = 0.1

Elementary analysis of 100 log(0.1 + count)

Blocks	0	5	10	20	30	eff	fit
1	18	-1	2	-85	65	-13	20
2	6	-14	11	28	-30	-22	11
3	-6	4	-105	44	62	-10	23
4	29	54	-55	-38	8	-60	-27
5	-34	-47	31	30	19	33	66
6	8	-3	56	16	-70	18	51
7	8	16	50	-8	-66	14	47
8	-25	-9	7	12	13	39	72
eff	83	52	-18	-35	-81	33	00
fit	110	85	5	-2	-48	00	-33

Regions with NE-SW shading correspond to 13 most positive comparison values, (trimean = +50); regions with NW-SE shading correspond to 13 most negative (trimean = -50).

The + 13		All residuals	The - 13
L		LLLL	LLL
-5	5	5	
-4	7	7	
-3	480	340	
-2	5	5	
-1		4	4
-0	39	61398	6
0	888	6884278	67
1		8616293	8932
2		98	9
3		10	0
4		4	
5		460	4
H		HH	

H	-34	H	-19h	H	-6
M	-3	H	+23h	M	12
H	+8			H	19
tri	-8			tri	+9
		midspread	43		

exhibit 13 (Cont'd)

B) FREE START = 0.25

Block	Elementary analysis of 100 log (0.25 + count) 0	5	10	20	30	eff	fit
1	30	12	11	50	9	-26	18
2	-9	-28	-8	1	42	- 7	37
3	19	- 3	-82	22	38	- 3	41
4	21	46	-33	-24	15	-52	- 8
5	-27	-39	30	22	10	25	69
6	9	- 1	39	3	51	16	60
7	7	15	41	-20	49	14	58
8	-12	- 3	1	3	3	32	76
eff	73	42	-19	-28	-69	44	00
fit	117	86	25	16	-25	00	-44

Trimeans of end 13's of comparison values = -44, +48

The + 13 L		All residuals L	The - 13
-5	0	01	1
-4		9	9
-3	39	93	
-2	47	7840	
-1	2	23	
-0	913	93189	9
0	97	9771333	33
1	55	925150	
2		122	12
3		0098	00
4		612	6
H			

H	-27	H	-16	H	33
M	1	H	+20	M	10
H	+ 7	dh	36	H	22
tri	- 6			tri	11
		midspread	36		

(Cont'd)

exhibit 13 (Cont'd)

C) FREE START = 1

Blocks	Elementary analysis of 100 log (1 + count)						
	0	5	10	20	30	eff	fit
1	13	- 3	- 8	-26	25	-10	52
2	1	14	- 1	11	2	-17	45
3	- 7	5	-39	21	24	- 9	53
4	- 3	21	-21	9	12	-27	35
5	-16	27	23	20	1	14	76
6	9	1	23	2	29	14	76
7	7	14	24	-17	26	13	75
8	- 8	7	2	4	- 5	20	82
eff	58	29	-14	-26	-47	62	00
fit	120	91	48	36	15	00	-62

Trimeans of end 13 comparison values = 10 and +10.

The + 13		All residuals	The - 13
-3		9	98
-2	7	76981	
-1	6	467	4
-0	189	887533211	2357
0	2977	112245779	114
1	12	1234	3
2	45	01133445	01

H	-9	H	-10h	H	-7
M	7	H	+12h	M	-2
H	11			H	4
tri	2			tri	-2
		midspread	21		

(Cont'd

exhibit 13 (Cont'd)

D) FREE START = 3

Blocks	Elementary analysis of 100 log (3 + count) by means						
	0	5	10	20	30	eff	fit
1	10	- 4	- 7	-11	13	- 7	77
2	- 1	-10	- 1	2	7	-13	71
3	- 7	1	-19	11	13	- 7	77
4	- 9	10	-10	2	10	-16	68
5	- 9	-18	14	12	- 1	7	91
6	11	0	11	- 6	-10	10	94
7	8	12	-11	-16	-16	10	94
8	- 1	10	- 1	0	- 6	12	96
eff	41	17	-10	-18	-30	84	00
fit	125	101	74	66	54	00	-84

Trimeans of end 13 comparison values = +3.4 and -3.1

The + 13		All residuals	The - 13
-1		9866	666
0			
-1		00	0
-0	9	997766	9766
-0	1	01111124	11
0	0	01	0
0	778	778	
1	001233	00001111122334	00
1			

H	-1	H	-7	H	-10
M	8			M	-6
H	11	H	10h	H	-1
tri	6			tri	-6
		midspread	17h		

(Cont'd)

exhibit 13 (cont'd.)

E) COMPARISON of DIAGOSTIC RESULTS'

Free Start	Diff of trimeans of residuals	Diff of trimeans of comp. values*	Ratio**	Diff. of effects	Clarity
0. 1	-17	100	-0.2	164 + 99 = 263	6.1
0.25	-17	92	-0.2	142 + 84 = 226	6.2
1	+ 4	20	+0.2	105 + 47 = 152	7.2
3	+12	6.5	+1.8	71 + 28 = 99	5.7

*Trimean for + 13 minus trimean for - 13.

**Ratio of difference of trimeans of residuals to difference of trimeans of comparative values.

F) COMPARISON of TREATMENT EFFECTS -- somewhat differently expressed

Free Start	Lead arsenate				
	0	5	10	202	30
	(as re-expressed values)				
0.1	83	52	-18	-35	-81
0.25	73	42	-19	-28	-69
1	58	29	-14	-26	-47
3	41	17	-10	-18	-30

G) COMPARISON of TREATMENT FITS -- converted back to counts

0.1	14.3	7.0	1.3	0.9	0.2
0.25	14.5	7.0	1.5	1.2	0.3
1	14.8	7.1	2.0	1.3	0.4
3	14.8	7.2	2.4	1.6	0.5

elementary analyses. Panel E hints that a free start between 0.25 and 1.00 might promote additivity--because the tilt of the diagnostic plot goes through zero -- and the clarity is high. If we had many more sets of data, we might convert this hint into something strong enough to trust.

Panel F compares the effects, differently expressed. We can see that there is close qualitative agreement about the results of the various analyses, but we are left uncertain about the quantitative situation. So we reconvert the treatment fits in panel G, finding only slight shifts in fitted counts from one free start to another.

The reconversion proceeds as in these examples:

- for a free start of 0.1 and 5 lbs. of lead arsenate, the fit is 85. This is 0.85 for log (0.1 + fitted count) , whence 0.1 + fitted count = 7.1 and fitted count = 7.0 .

- for a free start of 3 and 5 lbs of lead arsenate, the fit is 101. This is 1.01 for log (3 + fitted count), whence 3 + fitted count = 10.2 and fitted count = 7.2 .

We note that the mean values of the rule-of-thumb ratios are 0.25 and 0.29 for free starts of 0.1 and 0.25, rising to 0.42 for a free start of 1 and 0.58 for a free start of 3.

The reasonable conclusions seem to be:

- a wise choice of free start is not essential, but can be helpful.
- when it matters, the familiar technique of the diagnostic plot is likely to guide us effectively in the choice of a free start.

* * *

24L. Roots and logs without tables (optional)

One problem with the use of square roots is working in the desert. Given paper, pencil, and NO tables, what is the poor analyst to do? If he wants to use square roots, is his position helpless?

Some of us were taught how to "do square root" as part of arithmetic. Fewer of us remember how to do it. None of us should bother.

There is an easy way to obtain square-roots -- at least better and better approximations to square roots -- that is worth remembering. It depends on starting with a trial value and calculating

$$(\text{new trial}) = \frac{1}{2}\left\{(\text{trial}) + \frac{(\text{number})}{(\text{trial})}\right\}$$

as many times as are needed to make the answer settle down well enough for our purposes.

The largest count among the examples of 24G is 1936. The worst starting value I care to contemplate is 10. Taking this as the first "(trial)", we find, successively:

$$(\text{next trial}) = \frac{1}{2}\left(10 + \frac{1396}{10}\right) = \frac{1}{2}\,(10+139.6) = 75^-$$

$$(\text{next trial}) = \frac{1}{2}\left(75 + \frac{1396}{75}\right) = \frac{1}{2}\,(75+18.6) = 47^-$$

$$(\text{next trial}) = \frac{1}{2}\left(47 + \frac{1396}{47}\right) = 37.8 \text{ (error)}$$

$$(\text{next trial}) = \frac{1}{2}\left(37.8 + \frac{1396}{37.8}\right) = 37.366$$

$$(\text{next trial}) = \frac{1}{2}\left(37.366 + \frac{1396}{37.366}\right) = 37.363084$$

(The error in the third step was accidental, not intentional. It was left in to show how errors -- even major ones -- will be taken care of in this process -- so long as the arithmetic in one step -- the last step -- is correct).

The square root of 1936 is, to six decimals, 36.363083. We have thus come close -- far closer than we are ever likely to need to for practical data analysis.

A more reasonable start would have been 40, leading to

$$(\text{next trial}) = \frac{1}{2}\left(40 + \frac{1396}{40}\right) = \frac{1}{2}\,(40 + 34.9) = 37.45$$

$$(\text{next trial}) = \frac{1}{2}\left(37.45 + \frac{1396}{37.45}\right) = \frac{1}{2}\,(37.45 + 37.28) = 37.36$$

Here two steps already give us more accuracy than we will need.

So, some of you want to ask me, what about logs -- how can we use logs in the desert? It is easy enough to do quite well -- if you can do two times two times two ... far enough.

The first 10 powers of 2 are easy to write down:

$$2$$
$$2\times2 = 4$$
$$2\times4 = 8$$
$$2\times8 = 16$$
$$2\times16 = 32$$
$$2\times32 = 64$$
$$2\times64 = 128$$
$$2\times128 = 256$$
$$2\times256 = 512$$
$$2\times512 = 1024$$

Now $2^{10} = 1024$ is close to $1000 = 10^3$, so its logarithm (to base 10) is close to 3. Hence log 2 is close to $\frac{1}{10}(3) = 0.3$. This gives us the following approximate log table:

Number:	2	4	8	16	32	64	128	256	512	1024
Log:	.3	.6	.9	1.2	1.5	1.8	2.1	2.4	2.7	3.0

If we look at 2, 4, 8, 1.6, 3.2, 6.4, 1.28, 2.56, 5.12 and 10.24, and put these in order we have the approximate log table given in exhibit 14. As the last column of that table shows, the error in this calculation is negligible for any purpose for which 1-decimal logs are useful -- and for most uses of 2-decimal logs

Any one of us can reconstruct this table -- any time, anywhere -- and if one-decimal logarithms are good enough, as they are for so many purposes, we can just use this table as a critical table giving all numbers ≥ 1.28 but < 1.6 a log of 0.1, etc. (This shifts our logarithms down an average of about 0.05, but such shifts rarely disturb our analyses).

exhibit 14 of chapter 24 -- table

Approximate 1-decimal logarithm table

based on the powers of 2

Number	Approximate log	(*)
1	0.0	(.000)
1.28	0.1	(.107)
1.6	0.2	(.204)
2	0.3	(.301)
2.56	0.4	(.408)
3.2	0.5	(.505)
4	0.6	(.602)
5.12	0.7	(.709)
6.4	0.8	(.806)
8	0.9	(.903)

(*) Actual 3-decimal value of log (to base 10).

Finally, one more point. We are likely to use 1- and 2-decimal square roots and logarithms. We can save space, time and effort -- and make fewer errors -- if we shed the decimal points while doing the arithmetic. We have already often done this by multiplying our logs by 100. This both avoids confusion while doing the arithmetic and makes it easier for a viewer to do detective work on the numbers.

*****

24P. Problems for chapter 24

1. Ashby also gave counts for Salicornia Europea for larger quadrats in another location. Exhibit **15** has the data. Make a root-log-root interval display. Will roots or logs come closer to symmetry?

2. Exhibit **16** gives the numbers of deaths in three states -- Nevada, Maine and New York -- for each year from 1929 to 1936. What choice(s) of expression seem likely to balance spreads best?

3. Do the problem of exhibit 7.

4. Watson gives the number of lesions per half leaf found on the third and fourth leaves of tobacco plant inoculated with plant virus Hy. III. The data are given in exhibit **17**. What form of expressing counts leads to a reasonably satisfactory analysis?

5. Winsor and Clarke reported rather extensive results on the catch of plankton with different nets. Exhibit **18** presents estimated numbers of plankton of various kinds caught in five horizontal hauls in which one large net (75 cm. diameter) was towed at 30 meters depth while two small nets (12.7 cm. diameter) were towed one meter deeper and shallower respectively. What form of expression leads to a reasonable analysis? What are the results of that analysis?

6. In 1942 Beale reported on the results of certain insecticidal trials. Each of six blocks were divided into 12 plots, and each of six treatments were applied to a pair of plots, selected at random, in each block. (Since Beale's

exhibit 15 of chapter 24

Counts of Salicornia Europea for quadrats of 2500 cm^2

13	0	3445555567889
24	1	12334566789
to 32	2	11122356
to 26	3	00355689
18	4	7
17	5	8
16	6	169
	7	
13	8	0
	9	
12	1	01, 02, 04, 29, 35, 39
6	1	52, 71, 73
3	2	25, 32, 35

From: Table IV of E. Ashley 1935. The quantitative analysis of vegetation, Annals of Botany 49:779-802.

exhibit 16 of chapter 24

Numbers of deaths in three states, 1929 to 1936

	Nevada	Maine	New York
1929	1199	11,353	153,985
1930	1160	11,082	147,453
1931	1346	10,490	149,107
1932	1342	10,719	147,827
1933	1233	11,102	148,655
1934	1297	10,937	149,088
1935	1324	11,024	148,462
1936	1439	11,325	153,545

Source: Page 656 of the World Almanac and Book of Facts for 1944.

Their source: U. S. Bureau of the Census.

exhibit 17 of chapter 24

Watson's Counts of Virus Lesions on Half-leaves

	Third leaf		Fourth leaf	
Plant #	Right	Left	Right	Left
1	56	73	32	50
2	33	32	16	35
3	47	58	22	48
4	58	48	47	50
5	72	68	16	21
6	75	56	66	83
7	40	60	29	49
8	76	101	17	39
9	179	106	23	18
10	77	63	18	25

Source: M. A. Watson 1936. Factors affecting the amount of infection obtained by Aphis transmission of the virus Hy. III. Phil. Trans. Roy. Soc. London 226: 457-489, especially table IV on page 464.

exhibit 18 of chapter 24

Winsor and Clarke's data on the catch of plankton in nets of different sizes

(L = 1 large net; S = 2 small nets combined)

		counts -- some based on samples				
ul	Net	Calanus finmarchus cop. V	C. finmarchus adult	Metridia Halceus adult	Centropages typicus adult	Euthemist juv.
1	L	50500	21000	17400	55500	2490
	S	1263	617	751	2250	63
2	L	54500	15800	33000	25200	2800
	S	1560	401	1235	1260	111
3	L	71000	8950	79000	25500	2700
	S	2450	374	3520	1550	251
4	L	27800	4950	32000	19600	2420
	S	710	111	932	763	151
5	L	19800	3030	13200	33500	965
	S	642	120	324	1265	62

Source: C. P. Winsor and G. L. Clarke 1940. A statistical study of variation in the catch of plankton nets. J. Marine Research 3: 1-34, especially top of table VII at page 32.

report was methodological, we do not know what the six treatments were.) The response was the number of insects, of a specified kind, remaining on the plot. Exhibit 19 presents the data. What forms of expression are plausible? What are the results of applying them, using a resistant form of elementary analysis?

7. The World Almanac gives the annual deaths by certain causes as in exhibit 20. What expression leads to a simple analysis? Show the analysis.

8. The numbers of killings -- murders and non-negligent manslaughters in a number of large cities during 1961, 1964 and 1966 are as in exhibit 21. What expression analyzes best? What are the results?

9. Do the elementary analysis left to the reader in exhibit 9.

exhibit 19 of chapter 24

Beall's Counts of Numbers of Phlegemontius quinquemaculata

Treatment	Block					
	1	2	3	4	5	6
1	10,7	20,14	14,12	10,23	17,20	14,13
2	11,17	21,11	16,14	17,17	19,21	7,13
3	0,1	7,2	3,1	2,1	3,0	1,4
4	3,5	12,6	4,3	5,5	5,5	2,4
5	3,5	3,5	3,6	1,1	3,2	6,4
6	11,9	15,22	15,16	13,10	26,26	24,13

Source: G. Beall 1942. The transformation of data from entomological field experiments so that the analysis of variance becomes applicable. Biometrika 32: 242-262. Table 7 at page 246.

exhibit 20 of chapter 24

U.S. Deaths by Year and Cause

Year	Cause				
	Mo V.*	Falls	Burns	Firearms	Gases**
1957	38,702	20,545	6,269	2,369	1,143
58	36,981	18,248	7,291	2,172	1,187
59	37,910	18,774	6,898	2,258	1,141
1960	38,137	19,023	7,645	2,334	1,253
61	38,091	18,691	7,102	2,204	1,192
62	40,804	19,589	7,534	2,092	1,376
1963	43,564	19,335	8,172	2,263	1,489
64	47,700	18,981	7,319	2,275	1,360
65	49,183	19,984	7,900	2,344	1,526

* by motor vehicles
** poison gases

Sources: World Almanac and Book of Facts: - for 1963, page 304. - for 1966, page 299, - for 1968, page 895.

Their sources: National Office of Vital Statistics, National Vital Statistics Division, National Center for Health Statistics.

exhibit 21 of chapter 24

Killings in Large Cities

City	murders and non-negligent manslaughters		
	1961	1964	1966
New York	482	636	653
Chicago	365	398	510
Los Angeles	159	177	226
Detroit	136	125	214
Philadelphia	144	188	178
Baltimore	89	144	175
Washington	88	132	141
Cleveland	80	116	139
Dallas	99	149	120
New Orleans	52	82	113
St. Louis	77	120	106
Akron	16	12	17
Boston	26	52	58
Buffalo	19	21	24
Denver	32	33	39
Jersey City	11	17	16
Kansas City	49	48	59
Long Beach	10	17	20

Source: World Almanac and Book of Facts for: -1963, page 310. - 1966, page 307. - 1968, page 903.

Their source: Federal Bureau of Investigation.